TECNICAS ESTADISTICAS NO PARAMETRICAS

RUBEN ARDILA ARDILA, Psicólogo.

1—INTRODUCCION

El objeto del presente trabajo es presentar ante ustedes las técnicas estadísticas no paramétricas más usadas en la investigación psicológica. No pretendemos ser exhaustivos, y nos limitaremos a dar un resumen conciso de ellas, en la medida en que el tiempo lo permita. Remitimos al lector interesado a las obras de la bibliografía, donde podrá completar con detalle las ideas aquí expuestas.

Las técnicas no paramétricas de probar hipótesis son muy útiles en las ciencias psicológicas, y encajan admirablemente bien en los datos experimentales. A veces se llaman pruebas de distribución libre. Sus características principales son las siguientes:

1. No presuponen que los puntajes que se están analizando se encuentran en una población distribuída en cierta forma, por ejemplo, normalmente distribuída.

2. Muchas de estas pruebas pueden usarse con puntajes que no son exactos en sentido numérico sino que son simples rangos. Este es su principal mérito, y la razón por la cual son tan útiles en investigación psicológica.

3. Los cálculos son relativamente simples. Por esta razón permiten dedicar más tiempo a la formulación cuidadosa de los problemas y objetivos de la investigación.

4. Son útiles con pequeñas muestras, lo cual es importante en los casos raros, que no pueden tratarse con otras técnicas.

No es preciso insistir en este momento sobre la importancia de la *inferencia* en todas las ciencias, y de la inferencia estadística en las ciencias de la conducta. Inferir consiste en derivar una conclusión o probabilidad; su validez ha dado origen a muchos trabajos, tanto de filósofos de la ciencia como de matemáticos y otros especialistas. En estadística nos interesa inferir conclusiones sobre gran número de eventos con base en una pequeña parte de ellos; existen normas para tomar las *muestras*, sobre su tamaño, sobre las pruebas estadísticas que pueden utilizarse de acuerdo con el tamaño, etc.

Un problema típico es el siguiente: ¿Cómo sabemos que las diferencias observadas en dos muestras significan realmente que las poblaciones de las cuales se tomaron son diferentes? ¿Cómo podemos averiguar que no se deben solo al azar? ¿Cómo es posible saber si una muestra de puntajes procede de una población específica?

Estos son problemas de todos los días que el investigador tiene que solucionar. Con este fin se usan las técnicas paramétricas y las no paramétricas. Dedicaremos buena parte de este trabajo a la exposición de las ventajas y desventajas de las pruebas estadísticas no paramétricas, y a la descripción somera de cada prueba.

La palabra *parámetro* se usó primero en Psicología Experimental para designar las constantes que intervienen en las curvas de aprendizaje, y que cambian cuando se modifican las condiciones experimentales. Hoy damos este nombre a los valores de la población. Las técnicas estadísticas paramétricas aparecieron antes que las no paramétricas; pueden basarse en suponer que la población está normalmente distribuída, o que ambos grupos de puntajes se derivan de poblaciones que tienen la misma varianza o amplitud de puntajes.

Las técnicas de distribución libre o no paramétricas son más recientes. No requieren que se hagan presupuestos sobre los parámetros, y exigen por esto menos calificadores. Las técnicas paramétricas solo pueden usarse con puntajes que sean realmente numéricos; esto se debe al hecho de que los puntajes se suman, dividen y multiplican. Muchas técnicas no paramétricas se basan en el orden o rango de los puntajes, no en sus valores numéricos. Otras técnicas no paramétricas pueden ser útiles incluso en los casos en que es imposible ordenar los datos.

2—PROCEDIMIENTO ESTADISTICO USUAL

Antes de describir las técnicas no paramétricas veamos cuál es el procedimiento que se usa generalmente para confirmar o refutar una hipótesis, en relación con un grupo de datos. Los 6 pasos que llevan desde la formulación de la hipótesis de nulidad hasta la decisión acerca de su rechazo, deben usarse en todos los casos. En esta forma se logra un orden cuidadoso que puede ser seguido por otros investigadores. El procedimiento usual es el siguiente:

1. *Formular la hipótesis de nulidad* (Ho): Se desea rechazar, con el fin de aceptar la hipótesis alternativa (H1), que el investigador desea probar. La Ho de nulidad podría ser: No existe diferencia entre dos grupos, en lo referente a cierto rasgo determinado. Por ejemplo, dos grupos de ratas, uno bien alimentado y el otro hambriento no difieren en el número de callejones ciegos en el laberinto, ambos grupos cometen igual número de errores. La H1 sería: existe diferencia en el número de errores cometidos por ambos grupos. En algunos casos puede ser suficiente con constatar esa diferencia, en otros se desea probar cuál grupo comete más errores.

Esquemáticamente:

Ho : G 1 = G 2 (No hay diferencias).

H1 : G 1 = G 2 (Las hay).

$$\text{H1 puede ser} \begin{cases} G\ 1 < G\ 2 \\ G\ 1 > G\ 2 \end{cases}$$

2. *Elegir la prueba estadística:* Puede ser paramétrica o no paramétrica. Este punto será tratado detalladamente a lo largo del presente trabajo.

3. *Nivel de significación y tamaño de la muestra:* El nivel de significación debe enunciarse antes de recoger los datos, para saber qué detalles son pertinentes en la investigación y cuáles pueden descuidarse. Los dos niveles de significación más usuales son .05 y .01; se enuncian con la letra griega α (alfa). Se trata de saber si la probabilidad de ocurrencia de la hipótesis nula es igual o menor que alfa, con el fin de rechazarla y aceptar en cambio la hipótesis experimental, H1. Existen dos errores que pueden cometerse:

Error Tipo I: Rechazar la hipótesis nula siendo verdadera $= \alpha$.

Error Tipo II: aceptar la hipótesis nula siendo falsa $= \beta$.

Los dos tipos de errores están relacionados con el tamaño de la muestra. Generalmente se determinan α y N, el tamaño de la muestra, por anticipado; la probabilidad de cometer el error de tipo II (β) se determina después. Hay una relación inversa entre la posibilidad de cometer los dos tipos de errores. Para reducir ambos debemos aumentar el tamaño de la muestra.

El poder de la prueba es la probabilidad de rechazar la hipótesis nula cuando realmente es falsa.

Poder $= 1 - \beta$.

Si la hipótesis alterna, H1, tiene dirección, se usa una prueba unilateral (one tail). Es más poderosa que una prueba bilateral (two tails).

4. *Distribución de la muestra:* Es siempre teórica. Se considera como la frecuencia esperada. Permite predecir la probabilidad de ocurrencia de ciertos valores numéricos. Es la probabilidad asociada con la ocurrencia del evento.

5. *La región de rechazo:* Es una región de la distribución. La probabilidad asociada con cualquier valor en la región de rechazo es igual o menor que α (alfa).

6. *La decisión:* Rechazamos la hipótesis nula si la prueba estadística da un valor que se encuentra en la región de rechazo. Un valor significativo es aquel cuya probabilidad asociada de ocurrencia bajo la hipótesis nula es igual o menor que α (alfa).

3—ELECCION DE LA PRUEBA ESTADISTICA APROPIADA

Este es un importante problema que debe ser solucionado cuidadosamente. Buena parte del éxito de una investigación puede estar condicionada a la elección de una prueba estadística adecuada para ese caso concreto. Se considera que una prueba es buena si tiene una probabilidad pequeña de rechazar la hipótesis nula (Ho) siendo verdadera, pero una probabilidad grande de rechazarla siendo falsa. En general, deben tenerse en cuenta varios factores en la elección de la prueba estadística:

1. El poder de la prueba.
2. La forma en que se encontraron los puntajes de las muestras.
3. La naturaleza de la población de la cual se tomó la muestra.
4. La clase de medida o escala empleada.

Se relaciona directamente con el *modelo estadístico* que adoptemos. Es de todos conocida la importancia de los modelos en la Psicología contemporánea; uno de ellos, el de la cibernética, promete mostrarnos aspectos de nuestra ciencia que antes ignorábamos. Los modelos matemáticos, en general, han sido los más fructíferos en la explicación psicológica, y representan la rama de la Psicología que más interés despierta actualmente. Ciencia no es solo recolección de hechos sino también su explicación; en este sentido la matemática da un esquema lógico en el cual encajar los hechos observados empíricamente.

Volviendo a nuestro caso del modelo estadístico, sabemos que se construye con base en los presupuestos de la prueba. Las conclusiones tienen la forma siguiente: "Si el modelo usado fue correcto, y se han satisfecho las medidas requeridas, entonces...".

Sabemos siempre que cuanto menos presupuestos, más generales serán las conclusiones que se obtengan.

Las pruebas más poderosas tienen los presupuestos más grandes; las pruebas T y F, por ejemplo, tienen muchos presupuestos. Son las más adecuadas para rechazar la hipótesis nula cuando es falsa, siempre que los presupuestos sean válidos. Veremos más adelante cuáles son los presupuestos del modelo estadístico paramétrico, y en qué casos se cumplen en la investigación psicológica.

4—ESCALAS DE MEDIDA

En 1932 la Sociedad Británica para el Avance de la Ciencia reunió dos comisiones con el fin de estudiar lo relacionado con la teoría de las unidades de medida; la Sección A trabajó sobre Matemáticas y Física, la Sección J sobre Psicología. Se quiso saber si era posible en realidad medir la sensación humana, lo cual condujo a desacuerdo sobre el sentido del término medición, y finalmente a un estudio profundo sobre toda la teoría de las escalas.

Sabemos que la relación entre las cosas observadas y los números asignados debe ser tan directa, que actuando sobre los números aumentemos nuestro conocimiento de las cosas. La medición se presenta en diversas formas y las escalas pueden organizarse en clases definidas; cada una representa un nivel de medida diferente. Son cuatro (véase Cuadro número 1).

1. *Escala Nominal:* Se usan números u otros símbolos para clasificar un objeto, persona o cierta característica. Por ejemplo: numerar los jugadores de un equipo de fútbol, nombrar las enfermedades, etc. Es una escala muy primitiva, que solo permite relaciones de equivalencia; en otras palabras, el grupo se divide en subgrupos mutuamente excluyentes.

2. *Escala Ordinal:* Los objetos no solo son diferentes sino que están en cierta relación unos con otros. Además de la relación de equivalencia se usa la de ser "mayor que" (mejor que, preferido que, etc.). Puede tratarse de una escala parcialmente ordenada. Como ejemplos tenemos el estatus de las diversas clases sociales, la clasificación en la jerarquía militar y muchos más. Pueden realizarse diversas transformaciones matemáticas siempre que no se altere el orden relativo. Se utilizan aquí las pruebas de rangos. Debe haber un continuo en los puntajes observados. Las propiedades de la escala ordinal no son isomorfas con la aritmética. Esta escala es la que utilizan los psicólogos generalmente.

3. *Escala de Intervalo:* Tiene todas las propiedades de la escala ordinal, pero además se conocen las distancias entre dos números cualquiera. Se caracteriza por una unidad de medida constante y común. La razón de dos intervalos cualquiera es independiente de la unidad de medida y del punto cero; tanto la unidad de medida como el punto cero son arbitrarios. Como ejemplos pueden citarse las escalas de temperatura (Centígrado o Fahrenheit), las fechas, etc. Los psicólogos están muy interesados en lograr escalas de intervalo, pero esto no siempre es posible. Las diferencias en la escala son isomorfas con la estructura de la aritmética. Las operaciones matemáticas permitidas deben preservar el orden y las diferencias relativas entre los objetos. La escala de intervalo es la primera escala realmente cuantitativa.

4. *Escala de Razón:* Posee todas las propiedades de la escala de intervalo, y además un cero verdadero. La razón entre dos puntos cualquiera de la escala es independiente de la unidad de medida. Como ejemplo tenemos la escala de pesos: tiene cero verdadero, puede medirse en gramos o libras, la unidad de medida no influye. Esta es la escala usada en física. Es isomorfa con la estructura de la aritmética. Los números que se le asocian son números verdaderos; la unidad de medida es arbitraria. No se altera la relación entre dos números cuando todos los valores de la escala se multiplican por una constante positiva. Con esta escala puede usarse cualquier prueba estadística. Además de las usadas en la escala de intervalo, se utilizan el promedio geométrico y el coeficiente de variación, pues estos requieren que se conozca el cero verdadero.

5—PRUEBAS PARAMETRICAS Y NO PARAMETRICAS

Después de lo anterior, podemos ver qué características se requieren para que pueda utilizarse una técnica estadística con preferencia a otra. Un modelo estadístico *paramétrico* tiene ciertos presupuestos:

1. Las observaciones deben ser independientes.

2. Las observaciones deben provenir de poblaciones normalmente distribuídas.

3. Las observaciones deben tener alguna varianza.

4. Las variables incluídas deben medirse al menos en la escala de intervalo, de modo que permita operar en forma aritmética sobre los puntajes.

5. Debe existir homocedasticidad, o sea iguales varianzas en las poblaciones.

El modelo estadístico *no paramétrico*, por el contrario, no especifica las condiciones sobre los parámetros de la población de la cual se tomó la muestra. Presupone solamente que:

1. Las observaciones son independientes.

2. La variable que se estudia posee continuidad.

No requieren medidas muy poderosas, se aplican a la escala ordinal, e incluso algunas a la escala nominal.

Es preciso aclarar que una prueba paramétrica es más poderosa cuando se cumplen los presupuestos de su modelo estadístico, y las variables se miden al menos en la escala de intervalo.

Sabemos que por el concepto de poder-eficacia, aumentando el tamaño de la muestra podemos usar una prueba no paramétrica con el mismo poder de rechazar la hipótesis nula. Además los datos de las ciencias de la conducta rara vez cumplen las condiciones requeridas para usar las pruebas paramétricas. A esto se debe el puesto importante que ocupan las pruebas no paramétricas en psicología. Whitney (1948) afirma que para la mayor parte de las distribuciones de población, una prueba estadística no paramétrica es claramente superior en poder a una paramétrica. Sin embargo, las pruebas no paramétricas tienen ventajas y desventajas.

A) *Ventajas de las pruebas estadísticas no paramétricas:*

1. Las probabilidades que se obtienen con la mayoría de ellas son exactas, sin tener en cuenta la forma de la distribución de la población de la cual se tomó la muestra. Algunas pueden presuponer identidad de forma de dos o más distribuciones y otras presuponen distribuciones simétricas de las poblaciones. En ciertos casos presupone que la distribución es continua.

2. Si los tamaños de la muestra son tan pequeños como $N = 6$, no hay alternativa posible, debe usarse una prueba estadística no paramétrica a menos que se conozca exactamente la naturaleza de la distribución de la población.

3. Existen pruebas estadísticas no paramétricas para tratar muestras tomadas de observaciones de diferentes poblaciones. Ninguna prueba paramétrica puede tratar tales datos sin hacer presupuestos irreales.

4. Hay pruebas estadísticas no paramétricas para tratar datos que se dan en rangos, y datos cuyos puntajes numéricos tienen solo la fortaleza de rangos; o sea datos en los cuales no es posible realizar verdadera cuantificación ("A es más ansioso que B", pero no sabemos cuánto).

5. Los métodos no paramétricos pueden tratar datos que son simplemente de clasificación, que se miden solo en la escala nominal. Los métodos paramétricos no pueden hacerlo.

6. Las pruebas estadísticas no paramétricas son mucho más fáciles de aprender y de aplicar que las paramétricas.

B) *Desventajas de las pruebas estadísticas no paramétricas.*

1. Si se cumplen todos los presupuestos del modelo paramétrico, y si la me-

dida requiere cierta fortaleza, entonces las pruebas estadísticas no paramétricas constituyen un desperdicio de datos. El grado de desperdicio se mide por la relación poder - eficacia.

2. No existen métodos no paramétricos para probar interacciones en el modelo del análisis de varianza, a menos que se hagan ciertos presupuestos sobre aditividad. (En 1956 apareció un método, que se debe a K. V. Wilson, pero sus consecuencias no han sido aún completamente estudiadas).

6—DESCRIPCION DE CADA UNA DE LAS PRUEBAS

Ha llegado el momento de describir con cierto detalle las diferentes clases de pruebas estadísticas no paramétricas que se usan en la investigación psicológica. Es imposible profundizar en cada una de ellas, y los interesados pueden remitirse a la obra de S. Siegel, donde encontrarán ejemplos muy bien explicados (véase Cuadro número 2).

A) *Caso de una muestra.*

Hay pruebas estadísticas no paramétricas para probar una hipótesis que surge de una sola muestra. Nos dicen si una muestra particular procede de una población específica. Tomamos una muestra al azar y tratamos de probar la hipótesis de que viene de una población con una distribución específica. Puede usarse la prueba T que es paramétrica, para estudiar la diferencia entre el promedio observado (en la muestra) y el promedio esperado (en la población); presupone que las observaciones o puntajes en la muestra vienen de una población normalmente distribuída; y que se midan al menos en la escala de intervalos.

Puede no usarse la prueba T porque sus presupuestos no sean reales para los datos en cuestión, o por otras razones. Puede en cambio elegirse una prueba no paramétrica, de las usadas con una sola muestra. Hay cuatro pruebas de esta clase:

1. *Prueba Binomial:* Se usa con poblaciones de solo dos clases (por ejemplo hombre - mujer). No es posible esperar que una muestra al azar contenga la proporción exacta de P (una clase), y de Q (la otra) que se presenta en la población. La distribución binomial es la distribución de las muestras en las proporciones que podríamos observar en muestras al azar tomadas de una población de dos clases.

2. *Prueba X^2 de una muestra:* Se usa con dos o más categorías. Por ejemplo, si los individuos en una escala de actitudes pueden contestar que están "a favor", "indiferentes", o "en contra" de algo. Puede utilizarse para probar si existe una diferencia significativa entre el número *observado* de objetos o respuestas en cierta categoría, y el número *esperado* con base en la hipótesis nula.

3. *La prueba Kolmogorov - Smirnov de una muestra:* Trata sobre el acuerdo que se presenta entre la distribución de un grupo de valores de la muestra (puntajes observados), y una distribución teórica específica. Determina si los puntajes de la muestra pueden considerarse como provenientes de la población que contenga la distribución teórica. La prueba especifica la distribución de frecuencia acumulativa que podría ocurrir bajo la distribución teórica, y la compara con la distribución de frecuencia acumulativa observada.

4. *Prueba de "recorridos" de una muestra:* Un "recorrido" (run) es una sucesión de símbolos idénticos que son seguidos y precedidos por símbolos diferentes o por ningún símbolo. El número total de recorridos nos dice si la muestra es al azar o no. Si existen muchos puede haber fluctuaciones acíclicas de período corto influenciando los puntajes. Este método se basa en el orden de los eventos, y nos da información no

contenida en la frecuencia de los mismos.

Elección.

Tres de las pruebas anteriores son de bondad del ajuste (goodness of fit); la cuarta es de distribución al azar de una secuencia de eventos en la muestra. La elección entre las primeras tres pruebas puede determinarse por los factores siguientes:

a) Número de categorías en la medida.

b) Nivel de medida usado.

c) Tamaño de la muestra.

d) Poder de la prueba estadística.

La prueba binomial se usa cuando hay dos categorías en los datos; o cuando el tamaño de la muestra es tan pequeño que la prueba X^2 es inaplicable. Para usar la X^2 se requiere que los datos se encuentren en categorías discretas, y que las frecuencias esperadas sean suficientemente grandes. La prueba de Kolmogorov - Smirnov puede usarse cuando es posible presuponer que la variable que se considera tiene una distribución continua; esta prueba trata a los individuos separadamente, y no pierde información por causa de la agrupación, como hace la prueba X^2. La prueba de Kolmogorov-Smirnov es la más poderosa para verificar la "bondad del ajuste" (goodness of fit). La cuestión acerca de la eficacia de las pruebas basadas en recorridos (runs) tiene sentido solamente dentro del contexto de un problema específico.

B) *Dos muestras relacionadas.*

Se trata de saber si dos tratamientos son diferentes, o si uno es mejor que otro. Por ejemplo, comparar dos métodos de enseñanza; pero si se hace en dos grupos con diferentes capacidades o diferente motivación, los resultados pueden no tener mucho que ver con los dos métodos de enseñanza sino con las diferencias entre los grupos. Cuando estos se comparan, se pretende que estén bien pareados, siendo cada miembro tan semejante al miembro del grupo de control como sea posible. Es difícil igualar a los individuos, porque no se conocen todos los determinantes de la conducta. Lo más adecuado es que cada sujeto sea su propio control.

La prueba más comúnmente usada para analizar datos de dos muestras relacionadas es la prueba T para puntajes diferentes. Muchas veces es inaplicable, por diversas razones. Puede entonces preferirse una prueba no paramétrica. Existen 5 adecuadas para el caso de dos muestras relacionadas.

1. *Prueba de Mc Nemar de significación de cambios:* La persona es su propio control. Se trata de separar el "antes" y el "después" en el diseño experimental, y probar la efectividad del tratamiento realizado entre los dos.

2. *La prueba de signos* (Sign Test): Toma su nombre del hecho de usar los signos más (+) y menos (—), en lugar de medidas cuantitativas. Es útil cuando es imposible realizar medidas cuantitativas pero pueden lograrse rangos, uno con respecto a otro de los dos miembros de cada par. En este caso se desea establecer que las dos condiciones son diferentes; solo se presupone que hay una distribución continua en la variable que se considera. No se presupone nada sobre la forma de la distribución de las diferencias.

3. *Prueba de Wilcoxon de pares igualados, por rangos de signos:* Se usa cuando es importante la magnitud relativa, tanto como la dirección de las diferencias entre los pares. Es útil en psicología: generalmente podemos decir cuál miembro del par es "mayor", lo cual indica el signo de la diferencia entre sus miembros.

4. *Prueba de Walsh:* Se usa si es posible presuponer que los puntajes diferentes de dos muestras relacionadas provienen de poblaciones simétricas. No se requiere, sin embargo, que vengan de

poblaciones normales ni de la misma población. Se presume solo que son simétricas o sea que el promedio es una representación adecuada de la tendencia central y que es igual a la mediana.

5. *Prueba al azar por pares igualados:* Bajo ciertas condiciones esta es la prueba no paramétrica más poderosa. Puede ser usada siempre que la medición sea tan precisa que los valores de los puntajes tengan significado numérico.

Elección.

El caso de muestras relacionadas es aquel en cuyo diseño se usa el método de los pares igualados. Todas las pruebas aquí consideradas, menos la de Mc Nemar de significación de cambios, presuponen que la variable considerada tiene una distribución continua. El Mc Nemar se usa cuando una o varias de las condiciones que se están estudiando, se han medido solo en la escala nominal. La prueba de signo se usa si es posible la escala ordinal, pero el puntaje solo se ha medido en forma muy aproximada. La de Wilcoxon se prefiere cuando es posible hacer un rango con las diferencias observadas para los varios pares igualados. La prueba de Walsh cuando es posible presuponer que las poblaciones de las cuales se han tomado las muestras son tanto simétricas como continuas, cuando las medidas se realizan al menos en la escala de intervalos, y el tamaño de la muestra (N) es 15 o menos. Finalmente la prueba por azar de pares igualados se utiliza si el tamaño de la muestra no es muy grande, de modo que sea factible realizar los cómputos del caso.

C) *Caso de dos muestras independientes.*

No es necesario que las dos muestras a considerar posean el mismo tamaño. La técnica paramétrica usual es la prueba T, pero muchas veces no puede usarse y se prefiere una no paramétrica. Consideraremos 8 de tales pruebas.

1. *Prueba de Fisher de probabilidad exacta:* Es útil para datos discretos, sean nominales u ordinales, cuando las dos muestras independientes son de tamaño pequeño. Se usa cuando los puntajes de dos muestras independientes al azar caen en una u otra de dos clases mutuamente excluyentes. Fisher recomendaba esta prueba para todo tipo de datos dicotomizados, pero algunos han discutido su validez.

2. *Prueba X^2 para dos muestras independientes:* Puede usarse para determinar la significación de las diferencias entre dos grupos independientes, cuando los datos consisten en frecuencias que caen en categorías discretas. Es válido en la escala nominal. Su hipótesis reside en afirmar que los dos grupos difieren con respecto a alguna característica, y por lo tanto con respecto a la frecuencia relativa con la cual los miembros del grupo caen en las diversas categorías.

3. *Prueba de Mediana:* Se trata de probar si dos grupos independientes difieren en sus tendencias centrales. La prueba nos dice si es probable que los dos grupos independientes (no necesariamente del mismo tamaño) hayan sido tomados de poblaciones con la misma mediana. Los puntajes de los dos grupos deben estar al menos en la escala ordinal.

4. *Prueba U de Mann-Whitney:* Datos también en la escala ordinal. Se trata de probar si dos grupos independientes se han tomado de la misma población. Es una de las pruebas no paramétricas más poderosas; sirve como alternativa a la prueba T cuando se quieren evitar los presupuestos de ésta.

5. *Prueba de Kolmogorov - Smirnov de dos muestras:* Se desea probar si dos muestras independientes se han tomado de la misma población, o de poblaciones con la misma distribución. La prueba de dos colas es sensible a cualquier clase de diferencia en las distribuciones de las cuales se tomaron las dos muestras. La prueba de una cola puede usarse para decidir si los valores de la población de la cual se tomó una muestra son ma-

yores que los valores de la población de la cual se tomó la otra. Como la prueba debida a los mismos autores y que trata con una sola muestra, la presente prueba se refiere al acuerdo entre dos distribuciones acumulativas, en este caso los valores de las dos muestras.

6. *Prueba de Wald-Wolfowitz de "recorridos":* La hipótesis nula reside en afirmar que las dos muestras independientes han sido tomadas de la misma población. La hipótesis alterna (H1) es afirmar que difieren en algún aspecto. Con datos suficientemente amplios esta prueba rechaza la hipótesis nula si las poblaciones difieren en alguna forma: tendencia central, variabilidad, etc. Su ventaja es precisamente ésta, que trata con diferencias de cualquier clase, mientras que las otras pruebas se concretan solo a ciertos aspectos.

7. *Prueba de reacciones extremas de Moses:* Se usa cuando se espera que las condiciones experimentales afectarán a unos sujetos en una forma y a otros en la opuesta. En experimentos sobre "defensa perceptual" se ha comprobado que ciertas condiciones pueden producir en unos sujetos un estado de alerta extremado, y en otros un estado de inhibición también extremado. Esta prueba puede usarse asimismo en situaciones experimentales en las cuales un grupo es muy alto y el otro muy bajo. Es preciso que existan bases *a priori* para creer que las condiciones experimentales conducirán a puntajes extremos en cualquier dirección.

8. *Prueba al azar de dos muestras independientes:* Se quiere probar la significación de la diferencia entre los promedios de dos muestras pequeñas independientes. Requiere al menos medidas de intervalo. No es necesario presuponer distribución normal u homogeneidad de varianza en la población de que se trata.

Elección.

Las anteriores pruebas buscan saber si es factible que dos muestras independientes vengan de la misma población. Elegimos la prueba que sea más sensible a las diferencias que nos interesan; si se trata de diferencias de localización (tendencia central) debe elegirse, por ejemplo, la prueba de mediana (o la de Fisher si N es pequeño), la prueba U de Mann-Whitney, etc. La prueba de localización más poderosa es la de azar. Para cualquier clase de diferencia se prefiere generalmente la prueba de Kolmogorov - Smirnov.

D) *Caso de "K" muestras relacionadas.*

Se trata de probar la significación de la diferencia entre tres o más grupos relacionados. La hipótesis nula reside en afirmar que K muestras (siendo 3 o más) se han tomado de la misma población o de poblaciones idénticas. La prueba paramétrica que se usaría en este caso es el análisis de varianza o prueba F. Ella presupone que los puntajes se tomaron independientemente de poblaciones normalmente distribuídas, y que todas las poblaciones tienen la misma varianza, etc. Si no podemos presuponer esto para unos datos, o deseamos evitar presupuestos con el fin de aumentar la generalidad de los hallazgos, es posible usar una prueba no paramétrica, de las dos siguientes.

1. *Prueba Q de Cochran:* Es una ampliación de la prueba de Mc Nemar para dos muestras relacionadas; en este caso se consideran más de dos muestras. Se trata de saber si tales grupos igualados de frecuencias o proposiciones difieren entre sí en forma significativa. Puede usarse con datos en la escala nominal, o en la ordinal pero dicotomizada.

2. *Análisis de varianza por rangos en dos direcciones, de Friedman:* Los datos deben estar al menos en la escala ordinal. La hipótesis nula afirma que las K muestras se han tomado de la misma población. En cada muestra el número de casos será el mismo porque se han igualado.

Elección.

En este caso la elección es más fácil que en los anteriores. Se usa la prueba Q de Cochran cuando la medida de la variable que se estudia está en la escala nominal u ordinal dicotomizada. El análisis de varianza de Friedman puede usarse cuando los datos están al menos en la escala ordinal. Se conoce muy poco, realmente, sobre el poder de estas dos pruebas. La de Friedman parece ser muy poderosa, comparable a la prueba paramétrica más poderosa, la prueba F; debe usarse siempre que sea posible.

E) *Caso de "K" muestras independientes.*

Existen pruebas para medir la significación de las diferencias entre tres o más grupos o muestras independientes. La Ho es que las muestras independientes se han tomado de la misma población, o de K poblaciones idénticas. La prueba paramétrica adecuada en este caso sería el análisis de varianza de una dirección. Si no puede utilizarse en estos datos porque no satisfacen los presupuestos, se utiliza alguna de las tres pruebas no paramétricas que expondremos a continuación. Tienen la ventaja adicional de poder usarse con datos que están solo en la escala nominal (clasificación) o en la ordinal (rangos).

1. *La prueba X^2 para k muestras independientes:* Las frecuencias se presentan en categorías discretas. Se trata de determinar la significación de las diferencias entre k grupos independientes. Es una ampliación de la prueba X^2 para dos muestras independientes.

2. *Extensión de la prueba de mediana:* Los datos deben estar al menos en escala ordinal. Se trata de determinar si k grupos independientes (no necesariamente del mismo tamaño) se han tomado de la misma población o de poblaciones con medianas iguales.

3. *Análisis de varianza en una dirección por rangos, de Kruskal - Wallis:* La hipótesis nula reside en afirmar que las k muestras vienen de la misma población o de poblaciones idénticas. Esta prueba presupone que la variable que se estudia tiene una distribución continua. La medida debe estar al menos en la escala ordinal.

Elección.

Se prefiere el X^2 para muestras independientes cuando los datos se encuentran en frecuencias, y las variables están en la escala nominal o en categorías discretas en la escala ordinal. La segunda y tercera pruebas requieren que las medidas se encuentren al menos en la escala ordinal. Cuando se trata de elegir entre la segunda y tercera, se prefiere la última porque es más eficiente; usa más información; convierte los puntajes en rangos; y es más sensible a las diferencias entre k muestras de puntajes.

Antes de dejar este tema, recordemos que existen otras 4 pruebas no paramétricas para k muestras independientes; por ser muy especiales en su utilización no las expondremos en este lugar.

F) *Medidas no paramétricas de correlación.*

Como todos sabemos, la correlación puede ser el fin que se busca en una investigación psicológica, o ser solo uno de sus pasos. Son problemas diferentes descubrir la *existencia* de una asociación en la población, y medir el *grado* de asociación entre dos grupos de puntajes.

El coeficiente de correlación representa el grado de asociación. En pruebas paramétricas se usa el coeficiente de correlación de Pearson (r); requiere puntajes que representen la medida al menos en la escala de intervalos iguales; se debe presuponer que los puntajes vienen de una población normal. Las medidas de correlación no paramétricas son cinco:

1. *Coeficiente de Contingencia* (C): Mide la relación entre dos grupos de

atributos. Es útil cuando tenemos solo información en la escala nominal con respecto a uno o ambos grupos de atributos. No requiere que se presuponga continuidad para las varias categorías, ni que las series estén ordenadas.

2. *Coeficiente de correlación de rangos de Spearman* (r_s): Fue la primera técnica estadística basada en rangos que se desarrolló. Algunas veces se la representa como "ro". Requiere que ambas variables a comparar estén medidas al menos en la escala ordinal, o que los objetos o individuos que se estudien puedan ser arreglados en dos series ordenadas.

3. *Coeficiente de correlación de rangos de Kendall* (τ): Puede usarse con la misma clase de datos que el coeficiente de correlación de rangos de Spearman. Mide el grado de asociación entre dos grupos de rangos. Una de sus ventajas sobre el coeficiente anterior reside en que puede ser generalizado a coeficiente de correlación parcial.

4. *Coeficiente de correlación parcial de rango de Kendall* ($\tau xy.z$): Cuando se observa correlación entre dos variables, existe siempre la posibilidad de que sea debida a la asociación entre cada una de ellas y una tercera variable. Por ejemplo, la amplitud del vocabulario de un niño y su estatura parecen estar asociados; puede ser debido a que ambas variables se asocian con una tercera, la edad. Este problema y todos los similares, se tratan estadísticamente por medio de la correlación parcial. Para ello puede usarse la técnica de Kendall, siempre que los datos estén medidos al menos en escala ordinal. No se presume nada sobre el tamaño de la población de los puntajes.

5. *Coeficiente de concordancia de Kendall* (W): Mide la relación entre varios rangos de N objetos o individuos. Mientras que el r_s y el τ (tao) expresan el grado de asociación entre dos variables, medidas en rangos (o transformadas en rangos), el coeficiente W expresa el grado de asociación entre k de tales variables. Es útil para juzgar la confiabilidad entre varias pruebas, y para estudiar grupos de variables.

Elección.

Veamos cómo seleccionar entre las 5 técnicas no paramétricas de medir el grado de correlación de las variables en una muestra. El coeficiente de contingencia (C) debe usarse cuando los datos están en la escala nominal y no pueden ser ordenados en forma significativa. El Spearman tiene sobre los demás la ventaja de que es fácil de computar; el Kendall puede generalizarse a coeficiente de correlación parcial.

Tao (τ) y r_s son igualmente poderosos para probar la existencia de la relación. El coeficiente de correlación parcial de Kendall mide la relación entre dos variables, cuando una tercera (de la cual pueden ellas depender lógicamente) se mantiene constante. El coeficiente de concordancia, W, mide la asociación entre varios grupos de rangos de N entidades; es útil para determinar el acuerdo entre varios jueces o la asociación entre tres o más variables.

7—CONCLUSIONES

En el trabajo anterior hemos presentado las técnicas estadísticas no paramétricas en forma global. Hemos visto en qué difieren de las paramétricas, cuáles son sus ventajas y sus desventajas. Antes enumeramos los seis pasos que se distinguen en el procedimiento estadístico, siendo la decisión, el más importante. También dimos un vistazo de conjunto a las escalas de medida y a su importancia para la investigación psicológica.

Existen dos puntos sobre los cuales quiero llamar la atención de ustedes en este momento. El primero se refiere a la relación entre matemáticas y ciencias empíricas. Como anotó S. S. Stevens, el conocido psicólogo norteamericano, "La estatura de una ciencia se mide comúnmente por el uso que hace de las matemáticas". Pero la medición varía en grado, tipo y precisión, es un asunto muy relativo. Nuestro estudio comparativo de las técnicas paramétricas y no paramétricas lo muestra claramente.

Medir es comparar, establecer una relación de isomorfismo entre las cosas y los números que les asignamos, en forma tal que actuando sobre los números aumentemos nuestro conocimiento de las cosas. Se basa en el hecho de que las leyes lógico-matemáticas encajan en el universo. Son las leyes de nuestro pensamiento y al mismo tiempo se aplican a los fenómenos físicos. Bertrand Russell escribió:

"El universo es matemático no porque sepamos mucho sobre él sino porque sabemos muy poco; solo podemos descubrir sus relaciones matemáticas".

Quizás algún día un psicólogo u otro especialista en los fenómenos humanos tratará de explicar por qué las leyes lógico-matemáticas, que están en el observador, se aplican tan perfectamente al universo astronómico, al átomo y a la conducta de los organismos. Por qué existen ciertas relaciones que se nos imponen con un carácter de necesidad, por ejemplo los teoremas de una geometría, euclideana o no euclideana, si aceptamos sus axiomas; y por qué estos teoremas encajan en el mundo real, cuando fueron únicamente creación humana.

Este problema trasciende los límites que nos hemos fijado para el presente trabajo, pero está en el fondo del concepto de medición, que es el tema central que nos reúne aquí. Para solucionarlo necesitaríamos profundizar tanto en la filosofía de las matemáticas como en la psicología del pensamiento. La respuesta ambigua de que "el hombre refleja la lógica del cosmos" puede ser una gran verdad o un equívoco lingüístico. Esto queda aún por ser estudiado.

Otra incógnita de gran importancia es la siguiente: ¿muchos problemas relacionados con la medición psicológica se deberán acaso a que no estamos utilizando las matemáticas adecuadas a nuestro campo de trabajo? Kurt Lewin buscaba qué matemáticas podrían usarse con los fenómenos psicológicos y halló la topología, una rama muy nueva y que no parecía tener ninguna aplicación práctica; no profundizó suficientemente, pero sus modelos psicológicos han sido muy útiles y han aumentado grandemente nuestra comprensión del hombre.

Clark L. Hull, otro psicólogo, realizó un sistema axiomático que reunía los hechos encontrados en el laboratorio de Psicología y les daba una estructura matemática de postulados, teoremas y corolarios. Su admirable labor no ha tenido continuadores de su misma estatura.

Cabe preguntarnos si las ramas de la matemática moderna que no han encontrado ninguna aplicación en el mundo real, podrían ser utilizadas por la ciencia psicológica. Esto requiere un estudio exhaustivo de matemáticas modernas y de psicología científica que no ha sido aún realizado.

No es preciso insistir en la gran importancia teórica que tienen los dos puntos anteriormente tratados. Por eso me he permitido presentarlos a la consideración de ustedes.

BIBLIOGRAFIA

ANDERSON, R. L.; BANCROFT, T. A.: *Statistical Theory in Research,* New York: Mc-Graw Hill, 1952.

EDWARDS, A. L.: *Statistical Methods for the Behavioral Sciences,* New York: Rinehart, 1954.

FAVERGE, J. M.: *Méthodes stadistiques en Psychologie Appliquée,* Paris: P. U. F., 1960.

FERGUSON, G.: *Statistical Analysis in Psychology and Education.* New York: Mc-Graw Hill, 1959.

Garrett, H. E.: *Statistics in Psychology and Education*, New York: Mc Kay, 1962.

Guilford, J. P.: ***Fundamental Statistics in Psychology and Education***, New York: Mc Graw Hill, 1956.

Hoel, P. G.: ***Introduction to Mathematical Statistics***, New York: Wiley, 1962.

Lewis, D.: *Quantitative Methods in Psychology*, New York: Mc-Graw Hill, 1960.

Siegel, S.: *Nonparametric Statistics for the Behavioral Sciences*, New York: Mc-Graw Hill, 1956.

Stevens, S. S.: Mathematics, Measurement and Psychophysics, en S. S. Stevens (Ed.), *Handbook of Experimental Psychology*, New York: Wiley, 1960.

CUADRO NUMERO 1

ESCALAS DE MEDIDA

ESCALA	RELACIONES	ESTADISTICAS	PRUEBAS	EJEMPLOS
Nominal	1. Equivalencia	Modo Frecuencia Coeficiente de Contingencia.	No paramétricas.	Numerar los jugadores. Nombrar las enfermedades.
Ordinal	1. Equivalencia 2. Mayor que	Mediana Percentil Spearman r_s Kendall τ Kendall W	No paramétricas.	Clasificación militar. Dureza de minerales.
Intervalo	1. Equivalencia 2. Mayor que 3. Razón conocida de dos intervalos.	Promedio Desviación estándar. Correlación de Pearson Correlación múltiple.	No paramétricas. Paramétricas.	Temperatura Energía Fechas.
Razón	1. Equivalencia 2. Mayor que 3. Razón conocida de dos intervalos. 4. Razón conocida de dos valores cualesquiera de la escala.	Promedio geométrico Coeficiente de variación.	No paramétricas. Paramétricas.	Longitud Peso Densidad Resistencia.

CUADRO NUMERO 2

PRUEBAS ESTADISTICAS NO PARAMETRICAS

NIVEL DE MEDIDA	CASO DE UNA MUESTRA	CASO DE DOS MUESTRAS		CASO DE K MUESTRAS		MEDIDAS NO PARAMETRICAS DE CORRELACION
		RELACIONADAS	INDEPENDIENTES	RELACIONADAS	INDEPENDIENTES	
Nominal	Prueba binomial Prueba X^2 de una muestra.	Prueba de Mc Nemar de significación de cambios.	Prueba de Fisher de probabilidad exacta. Prueba X^2 para dos muestras independientes.	Prueba Q de Cochram.	Prueba X^2 para K muestras independientes.	Coeficiente de contingencia: C.
Ordinal	Prueba de Kolmogorov - Smirnov de una muestra. Prueba de recorridos, de una muestra.	Prueba de Signos Prueba de Wilcoxon.	Prueba de Mediana. Prueba U de Mann Whitney. Prueba de Kolmogorov - Smirnov de dos muestras. Prueba de recorridos de Wald Wolfowitz. Prueba de reacciones extremas de Moses.	Análisis de varianza en dos direcciones de Friedman.	Extensión de la prueba de mediana. Análisis de varianza en una dirección, de Kruskal Wallis.	Coeficiente de correlación de rangos, de Spearman: rs. Coeficiente de correlación de rangos de Kendall: τ Coeficiente de correlación parcial de rangos de Kendall: $\tau xy.z$. Coeficiente de concordancia de Kendall: W.
Intervalo	Prueba de Walsh Prueba al azar por pares igualados.	Prueba al azar de dos muestras independientes.				

(Cuadro tomado y traducido de SIEGEL, op. cit).